Protein thermal stability: hydrogen bonds or internal packing?

Gerhard Vogt and Patrick Argos

Thermally stable proteins are of interest for several reasons. They can be used to improve the efficiency of many industrial processes and provide insight into the general mechanisms of protein folding and stabilization. Comparison of tertiary structural properties of several protein families with members of different thermostability should help to delineate the role of individual factors in achieving stability at high temperature. In this work, 16 protein families with at least one known thermophilic and one known mesophilic tertiary structure were examined for the number and type of hydrogen bonds and salt links, polar surface composition, internal cavities and packing densities, and secondary structural composition. The results show a consistent increase in the number of hydrogen bonds and in polar surface area fraction with increased thermostability.

Address: European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 102209, D-69012 Heidelberg, Germany.

Correspondence: Patrick Argos
E-mail: argos@embl-heidelberg.de

Electronic identifier: 1359-0278-002-S0040

Folding & Design 01 Aug 1997, 2:S40–S46

Introduction

If the human body temperature rises only a few degrees above 40°C, death is imminent. By contrast, there are bacteria living happily in the boiling water of hot springs or in the neighborhood of submarine volcanoes. Almost 40 years after the first 3D protein structure was determined, it is still a matter of debate which structural factors are the main determinants of increased stability in thermostable proteins. A major reason for this is the small difference in free energy between the folded and unfolded states of a protein and the large number of partly antagonistic factors, which are difficult to assess on an energetic scale. Nevertheless, thermal stability is of central importance in both science and industry. Increased thermostability of enzymes is a means to increased productivity. Chemical reactions run at higher speed at increased temperature, and the final purification process can be simplified if unwanted reaction products can be more easily eliminated at the high temperature. Thermostable proteins provide a means to better understand protein stability and folding, essential in engineering, design, and structure prediction of biomolecules. Most thermostable proteins have close mesostable relatives that share a basic mainchain fold and display sequence similarity. Comparing the tertiary structures of families with such constituents allows a systematic survey to focus on individual differences and therefore gain knowledge through consistent observations.

The issue of thermostability has been addressed numerous times during the past few decades as evidenced by several review articles [1–4]. While early reports were limited to the examination of protein sequence information [5,6], the availability of an increasing number of tertiary structures has allowed a more detailed and inclusive investigation. Querol *et al.* [3] recently listed at least 15 different physical and chemical reasons that researchers have reported over the years to explain enhanced thermostabilization after an examination of 3D structures before and after residue mutation. The explanations given are listed in Table 1 along with the number of citations for each [3]. These explanations point to two dominant causes: increased numbers of hydrogen bonds (and salt links) and better internal van der Waals' packing. Table 1 indicates that nearly 90% of the literature citations of various etiologies can be assigned (in whole or significant part) to either of the two phenomena, thus providing a focus for this theoretical investigation of protein tertiary structures. Families containing both thermostable and

Table 1

Physicochemical explanations for enhanced thermostability as reported in the literature.

Explanation type	Citation count
Better hydrogen bonding*	18
Better hydrophobic internal packing†	16
Enhanced secondary structure propensity*	12
Helix dipole stabilization*	10
Argos' replacements*†	10
Removal of residues sensitive to oxidation or deamination	10
Burying hydrophobic accessible area*†	7
Improved electrostatic interactions*	6
Strengthening intersubunit association*†	6
Decreased backbone strain*	5
Salt bridge optimization*	4
Better van der Waals' interactions†	3
Better affinity for calcium	2
Improved enthalpy upon substitution	1

The numbers of citations Querol *et al.* [3] found in their literature survey. *Explanations for thermal stability increase that can be reasonably attributed (in whole or at least in significant part) to increased hydrogen bonding (to water or protein atoms) and salt links. †Explanations that are related to better packing.

Table 2

Protein families containing at least one thermophile.

Family no.	Family name	PDB ID / average host living temperatures (°C) / structure resolution (Å)
1	Malate dehydrogenase	4mdh/37/2.5, 1bmd/72.5/1.9
2	Glycosyltransferase	1cdg/35/2.0, 1cgt/35/2.0, 1cyg/52.5/2.5, 1ciu/60/2.3
3	Glyceraldehyde-3-phosphate dehydrogenase	4gpd/20/2.8, 3gpd/37/3.5, 1gad/37/1.8, 1gd1/52.5/1.8, 1cer/71/2.5, 1hdg/82.5/2.5
4	Lactate dehydrogenase	6ldh/20/2.0, 1llc/35/3.0, 5ldh/37/2.7, 9ldb/37/2.2, 1lld/39/2.0, 1ldn/52.5/2.5
5	Thermolysin	1npc/30/2.0, 1lnf/52.5/1.7
6	Ribonuclease H	2rn2/37/1.48, 1ril/72.5/2.8
7	Subtilisin	1st3/30/1.4, 1sup/35/1.6, 1sca/42.5/2.0, 1thm/60/1.37
8	Ferrodoxin	1fca/28/1.8, 1fdx/37/2.0, 2fxb/52.5/2.3
9	Superoxide dismutase	3sdp/27.5/2.1, 1abm/37/2.2, 1isa/37/1.8, 1ids/37/2.0, 3mds/72.5/1.8
10	Phosphofructokinase	2pfk/37/2.4, 3pfk/52.5/2.4
11	Phosphoglycerate kinase	3pgk/27.5/2.5, 1php/52.5/1.65
12	Triose phosphate isomerase	1ypi/27.5/1.9, 1hti/37/2.8, 1tim/37/2.5, 1tpe/41/2.1, 1btm/52.5/2.8
13	Rubredoxin	6rxn/35.5/1.5, 1rdg/35.5/1.4, 8rxn/35.5/1.0, 5rxn/37/1.2, 1caa/110/1.8
14	Hydrolase	1ino/37/2.2, 2prd/72.5/2.0
15	Glycosyltransferase	2exo/30/1.8, 1xyz/60/1.4
16	Reductase	1lpf/27.5/2.8, 1lvl/27.5/2.45, 3lad/37/2.2, 1edb/52.5/2.6

mesostable proteins (see Table 2) were extracted from the current version of the PDB database [7], which contains files of atomic coordinates, and different properties were then examined for their correlation with enhanced thermostability within each family. The number of hydrogen bonds and salt links showed a clear increase with increased thermostability for >80% of the families, while the protein packing density and the number and fractional volume of protein cavities showed a less obvious consistency, albeit with some trend to lower cavity volume for relatively large cavities with increased thermostability.

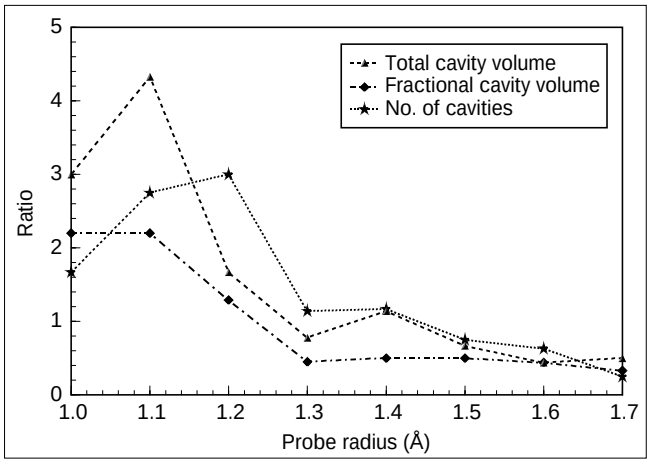
Results

Packing density and protein cavities

The number and size of cavities in a protein strongly depend on the diameter of the probe sphere used to calculate them [8]. An empty volume within a protein is considered as a cavity if the probe, which scans the solvent-accessible surface, cannot enter the protein from the exterior. If the probe radius is only marginally increased, narrow internal cavities just wide enough to fit the probe disappear completely or are at least split into several smaller cavities. The volume of a probe with a radius of 1.3 Å is ~10 Å^3, which is the minimum loss in cavity volume if one cavity is removed. In the more common case of irregularly shaped cavities, the cavity volume loss due to an increase of the probe radius of just 0.1 Å can be considerably greater.

Figure 1 shows the change in the ratio between the number of families with positive (increase in a property) and those with negative trends in cavity number and total and fractional cavity volume with increasing thermostability. All three ratios are above 1.0 for small probe sizes (≤ 1.2 Å), i.e. the number of very small cavities increases with increased thermostability. All ratios tend to decrease with increasing probe radius. The fractional cavity volume reaches a near stationary value of 0.45, i.e. a decrease in value for 11 of 16 families with increasing thermal stability when calculations are based on a probe radius of 1.3 Å or larger. Hubbard and Argos [8] proposed the use of probe radii between 1.2 and 1.3 Å after an extensive survey of

Figure 1



Plot of the probe radius in Å used to determine the cavities and the ratio of the number of families showing an increase of the specified property to the number of families displaying a decrease in the indicated cavity characteristic with increasing temperature stability.

Table 3

Change in fractional cavity volume, number of cavities, and packing density per 10°C rise in thermostability.

Family no.	Fractional cavity volume	No. of cavities	Packing density
1	0.11	11	0.30
2	1.24	15	0.21
3	–0.51	32	–2.08
4	1.13	4	–0.60
5	–1.42	13	3.43
6	0.72	5	–1.98
7	–0.69	–4	0.75
8	–0.50	0	4.30
9	0.72	101	1.43
10	–1.88	–38	4.52
11	–2.33	–24	–3.05
12	–1.47	–52	–1.06
13	–0.17	–1	0.21
14	–0.05	–5	0.16
15	–1.58	43	–1.09
16	–0.03	–41	–3.73
Mean	–0.97	28	1.87
Ratio	5+/11–	8+/7–	9+/7
Z-score	1.3	0	0.4

The fractional cavity volume is defined as the ratio of the total cavity volume to the volume contained within the contact surface of the protein. The changes given here with increasing thermal stability (for a 10°C rise) for cavity volume and packing density are given in % (i.e. multiplied by 100). Ratio indicates the number of families that show decreasing (–) or increasing (+) parameters with increasing host living temperature. The probe radius used for calculation of cavities was 1.3 Å. The mean is given over all families showing the most consistent trend. *Z*-scores for the ratios are calculated as described in Methods.

Table 4

Change in the number of hydrogen bonds and ion pairs per 10°C rise in thermostability.

Family no.	No. H-bonds per chain	No. H-bonds per residue	No. ion pairs per chain	No. ion pairs per residue
1	13.1	0.045	–0.56	–0.0015
2	–4.7	–0.005	1.75	0.0026
3	19.4	0.059	–0.07	–0.0002
4	4.4	0.026	–1.45	–0.0040
5	5.3	0.015	–0.89	–0.0029
6	–7.0	–0.033	3.10	0.0212
7	0.8	–0.007	1.61	0.0057
8	7.6	0.029	0.80	0.0043
9	11.8	0.049	0.52	0.0021
10	12.3	0.004	7.10	0.0220
11	60.4	0.164	3.20	0.0082
12	6.3	0.019	0.53	0.0019
13	1.0	0.017	0.10	0.0019
14	4.8	0.029	0.28	0.0017
15	4.8	0.007	–0.50	–0.0019
16	–4.9	0.005	1.81	0.0045
Mean change/ 10°C	11.7	0.036	1.89	0.0069
Ratio	13+/3–	13+/3–	11+/5–	11+/5–
Z-score	2.6	2.6	1.6	1.6

Values are given per polypeptide chain in each molecule to remove any bias from subunited complexes. Means are given only for those families showing a positive trend. *Z*-scores for the ratios are calculated as described in Methods.

cavities in globular proteins [8]. All subsequent statistics given in this work are based on the 1.3 Å probe. It must nonetheless be emphasized that the cavity trends reverse between 1.2 and 1.3 Å probe radii with increased thermal stability (see Discussion).

Table 3 lists the change in fractional cavity volume and in the number of cavities with increasing thermal stability for all 16 examined protein families. The change in packing density is also given. The average listed is that for the families with the most consistent trend. The number of cavities increases and decreases in about the same number of families. The results for the packing density also show no strong trend. The fractional cavity volume, however, decreases in 11 of 16 families (69%) for a 1.3 Å probe.

Hydrogen bonds and salt links

Table 4 lists the changes in the number of hydrogen bonds and salt links per 10°C rise in thermostability for all families. Scores are given as 'per chain' and 'per residue'. The ion pair counts refer to those not considered as hydrogen bonds. Both counts show an increase with increased thermostability for the majority of the families. This is especially true for the number of hydrogen bonds, which increases for 13 of the 16 families (81%) with an average gain for the families displaying an increase of 11.7 hydrogen bonds per chain (0.036 per residue) per 10°C rise in thermostability. A less dramatic trend (11 of 16 families, or 69%) is found for the number of salt links which increase by 1.89 per chain or 0.0069 per residue.

Protein surface composition

Polar atoms can act as hydrogen-bonding partners and are therefore energetically favorable at the water protein surface. Table 5 lists the fractional surface contribution of different atom types. The apolar carbon and sulfur fractional surfaces show a clear decrease of 0.81% and 0.17% of the total surface per 10°C rise in thermostability. The polar nitrogen atoms increase by 0.91% per 10°C rise in thermostability for 10 of the 16 families. Overall, the polar surface (nitrogen and oxygen) increases in 13 of the 16 families with an average change of 0.86% per 10°C rise in thermostability. In Table 6, the mainchain and sidechain polar atomic contributions are differentiated. Interestingly, the fraction of mainchain atoms generally decreases both for polar and apolar atoms. For the

Table 5

The change in fraction (in %) of total protein solvent-accessible surface area per 10°C difference in host stability temperature contributed by given atom types.

Atom types	Mean change/10°C	Ratio	*Z*-score
N	0.91	10+/6–	1.0
O	0.27	8+/8–	0.1
C	–0.81	3+/13–	2.5
S	–0.17	4+/10–	1.1
Net N+O	0.86	13+/3–	2.5

The fractional changes were calculated for each family according to equation 1 for a 10°C rise in thermal stability temperature. Ratio indicates the number of families that show decreasing (–) or increasing (+) surface for given atom types with increasing host living temperature. *Z*-scores for the ratios are calculated as described in Methods.

Table 6

Change in fractional surface contribution (in %) of different atom types by their affiliation with the mainchain or sidechains.

	Mainchain			Sidechain		
	Mean change/ 10°C	Ratio	*Z*-score	Mean change/ 10°C	Ratio	*Z*-score
C	–0.47	6+/10–	1.9	–0.76	4+/12–	1.9
N	–0.16	5+/11–	1.5	0.86	11+/5–	1.5
O	–0.38	6+/10–	0.9	0.83	9+/7–	0.3
S	–	–	–	–0.17	5+/10–	1.1
N+O	–0.49	6+/10–	0.9	0.93	14+/2–	3.1

The change in fractional surface per 10°C rise in host temperature is given as a mean over those families showing the most consistent trend. *Z*-scores for the ratios are calculated as described in Methods.

sidechain atoms, the polar trends are strongly reversed. The contribution of polar sidechain atoms rises by 0.93% of the total surface per 10°C rise in thermostability for 14 of the 16 families examined. Table 7 shows which residue types mostly contribute to the increase in polar surface (Arg and Glu) and which are preferentially removed to yield decreased hydrophobic surface (Ala, Cys, Phe and Val). Proline, however, contributes to the hydrophobic surface.

Secondary structural properties

Table 8 shows the changes in secondary structural assignments among aligned sequence pairs. The numbers are weighted by the square of the temperature and relate the mean net increase (or decrease) of residues over all families adopting given secondary structural types for an increase in thermostability of 100°C rise in temperature. There is a clear trend to helix and away from turn and coil in thermally stable structures. Table 9 shows the top five residue substitutions that transform backbone conformations to helix or strand.

Discussion

The experimental approach to thermal stability often uses site-directed mutagenesis to modify existing proteins. It is, however, applied in any one investigation to only one or a few protein types. Given all the reasons cited in the literature for increased thermal stability in mutated tertiary structures (Table 1), two major explanations appear: better internal packing and an increase in hydrogen bonds and electrostatic interactions. With the increasing number of tertiary structures in the PDB database, it is now possible to compare such information over many different protein families, although the dataset is still limited. In this work, we investigated 16 protein families with 56 individual constituents taken from the PDB in October 1996.

The packing density considers all the space contained within the protein's surface that is not filled with hard sphere protein atoms of appropriate van der Waals' radius [9]. In nine of 16 families (56%) investigated here there is an increase in packing density with increased thermostability, while in seven the opposite trend is observed. Clearly global packing density is not a significant stabilizing factor (*Z*-score 0.4) given present structural knowledge. The same is true for the number of cavities, which increases in eight families and decreases

Table 7

Number of families with increasing (+) or decreasing (–) accessible surface area, contributed by different sidechain (SC) atom types of a given residue type, with increasing temperature of thermal stability.

Residue type	SC_O	SC_N	SC_C and/or SC_S
Ala	–	–	6+/10–
Arg	–	11+/4–	11+/–4
Asn	8+/8–	8+/8–	7+/9–
Asp	6+/10–	–	9+/7–
Cys	–	–	3+/12–
Gln	9+/7–	6+/10–	8+/8–
Glu	11+/4–	–	11+/5–
Gly	–	–	–
His	–	6+/8–	7+/8–
Ile	–	–	9+/7–
Leu	–	–	7+/9–
Lys	–	8+/8–	8+/8–
Met	–	–	7+/9–
Phe	–	–	5+/11–
Pro	–	–	11+/5–
Ser	4+/12–	–	6+/10–
Thr	4+/12–	–	6+/10–
Trp	–	3+/11–	6+/9–
Tyr	9+/7–	–	10+/6–
Val	–	–	6+/10–

Table 8

Migration between different secondary structural types.

SS-type (S)	X→S	S→X	Net sum
H	30.95	16.56	14.39
E	23.46	21.07	2.39
G	20.42	11.65	8.77
T	60.68	70.05	–9.37
B	5.71	7.34	–1.63
C	47.94	62.49	–14.55

SS-type: H, helix; E, extended state; G, 3_{10}-helix; T, reverse turn; B, isolated β-bridge; and C, coil. X stands for any other structural assignment. X→S are all changes from any secondary structural type to structure type S and vice versa in the aligned sequences. Net sum refers to the net movement toward (+) or away (–) from the secondary structural type as the thermal stability increases. Values are given per 100°C rise in temperature and averaged over all families.

in seven as thermostability increases. However, the fractional cavity volume decreases in 11 of 16 families (69%, Z-score 1.3) with increased thermostability when a probe radius of 1.3 Å (or greater) is used (Fig. 1). Almost 1% of the total protein volume is converted from cavity to non-cavity for every 10°C rise in thermostability. The trend is reversed, however, when a probe radius of 1.2 Å (or less) is utilized (Fig. 1). The results from the smaller probes would consider smaller cavities and thus approach those of the packing density calculations. Hubbard and Argos [8] suggest the use of probes with radii between 1.2 and 1.3 Å, which are sufficiently large to avoid errors in structure determination as they yield consistent and spatially equivalent cavities across various tertiary structures determined by different researchers for the same protein under various crystallization, data collection, and structure refinement procedures. Thus, improvements in packing density or reduction in volume of relatively large cavities are not as significant (if at all) as increased hydrogen bonds in enhancing protein thermal stability.

Internal hydrogen bonds and salt bridges show a clear increase with increased thermostability (Table 4, respective *Z*-scores of 2.6 and 1.6). The increase in hydrogen bonds per chain and per 10°C rise in thermophilic stability was found to be 10.6 when both hydrogen bonding partners were buried, i.e. each had <10 $Å^2$ of solvent-exposed surface. This represents 91% of the entire hydrogen bond increase. A single hydrogen bond lowers the protein's free energy by ~0.5 to 2 kcal mol^{-1} [10]. The gain in free energy for ion pairs is estimated to range from ~0.4 to 1.0 kcal mol^{-1} [11,12]. Considering the average addition of 13 such bonds or links per 10°C rise in thermostability observed here and the experimental value of 0.5 kcal mol^{-1} per hydrogen bond [10], the free energy of the folded state would be reduced by 26 kcal mol^{-1} for a 40°C rise in thermostability. This is close to the 20 kcal mol^{-1} expected and approximately observed [13].

Table 9

Top five residue exchanges (mesophiles→thermophiles) observed in secondary structural substitutions to helix or strand in thermophiles.

To helix		To strand	
Substitution	Frequency	Substitution	Frequency
Gly→Ala	9.3	Val→Ile	9.5
Asp→Gln	8.0	Gly→Ala	6.5
Phe→Leu	7.7	Ser→Thr	6.0
Val→Ala	7.5	Lys→Arg	5.4
Lys→Arg	7.1	Leu→Ile	4.5

The total number of substitutions observed for helices and strands is 275 and 228, respectively, over all families.

It is also noteworthy that in the three families in which hydrogen bonds are not increased in the thermophiles, there is an above average increase in salt links. The results are also consistent among the families despite various possible X-ray crystallographic errors, structure resolutions and refinement conditions as well as different folds and functions across the proteins families examined.

The fractional surface contribution of polar atoms (nitrogens and oxygens) in thermally stable proteins is increased for 13 of 16 families (*Z*-score 2.5) if all atoms are considered (Table 5) and for 14 of 16 families (*Z*-score 3.1) if only sidechain atoms are counted (Table 6). Clearly, an increase of hydrogen bonding density to water at the protein surface aids in thermal stabilization. There is also a trend, albeit weaker, to increased secondary structure (especially helix) and decreased turn and coil conformations in thermally stable proteins (Table 8). This result is consistent with the observed increase in internal hydrogen bonds. Among the principal residue substitutions associated with increased secondary structure are Gly→Ala and Lys→Arg, both consistent with earlier observations by Argos *et al.* [6] and Menéndez-Arias and Argos [14]. They also constitute favorable substitutions to increase mainchain hydrogen bonds for both strand and helical conformations. Lys→Arg is also important in increasing the nitrogen polar surface. Further, arginine sidechain nitrogens are often involved in increased internal hydrogen bonds.

There seem to be a very large number of possible explanations for increased thermostability in nature. A closer look shows, however, that most of them are related to two main causes: hydrogen bonds and improved atomic packing. This work supports the former more strongly. It is noteworthy that in five recent reports of tertiary structures for different hyperthermophilic proteins [15–19], researchers in comparing with mesophilic counterparts suggested increased hydrogen bonding and/or ion pair networks as dominant factors in thermal stabilization.

Materials and methods

Choice of protein families

The protein families used in this work were chosen from the PDB protein atomic coordinate database [7]. First, a list of thermostable proteins was generated by searching all entries containing X-ray structures of wild-type proteins with one of the words 'therm', 'sulfolobus' or 'pyroc' in the names of the source species. The PDB contained 78 such structures in October 1996. If two or more structures were available for the same protein, the one with higher resolution, lower number of ligands, and later submission date was chosen according to the given order. The primary sequences of the resulting 35 structures were then compared against all primary structures in the PDB with a length not shorter than 66% and not longer than 150% of the length of the thermostable protein and with a sequence identity of at least 35% after alignment to assure relationship [20]. These sequences were clustered and identical proteins were removed following the criteria described above. In the final step, a structural superposition of the C_α atoms [9,21] amongst all family member pairs was determined; a visual inspection assured similarity of the 3D fold. When necessary, symmetry-related subunits of the molecules were generated using the information in the PDB files or, when not available, from the original publication of a given structure. For some proteins a few sidechain atoms were missing; the computer program ICM [9,21] was used to generate them in their energetically favorable positions.

A total of 56 proteins belonging to 16 protein families with at least one thermostable and one mesostable protein were found (Table 2). For these proteins, temperatures representing those of the host's optimal growth or normal living environment were extracted from the literature [2,14,22–25]. If a host had a habitat in a range of temperatures, the temperature average was taken. Host temperatures provided a consistent comparison of thermal stability. Experimental assays of specific enzyme stability and activity are not always available, especially for mesophiles, and are performed under various chemical conditions. The proteins represent a variety of folding architectures with various functions including mixed strand/helix topologies, symmetrical β-barrels, mostly extended structures and those binding iron-sulfur cages.

Evaluation of results

Structural property comparisons are performed only amongst the members of an individual family; the results are then averaged over all families to elicit trends. The score (S_j) associated with the jth family for a given property change from mesophile to thermophile was calculated per 10°C rise in thermal stability as:

$$S_j = \frac{\sum_{i=2}^{n}\sum_{k=1}^{i-1}(T_i - T_k)^2 * \left[\frac{(X_i - X_k)}{T_i - T_k}\right] * 10}{\sum_{i=2}^{n}\sum_{k=1}^{i-1}(T_i - T_k)^2} \qquad (1)$$

where X_i represents the property value of the ith family member, T_i is the corresponding thermal stability temperature assigned to the host (Table 2) with T_i always taken as greater than T_k, and n is the number of constituents in family j. The squared weight was adopted to emphasize the results from pairs with large environmental temperature differences where characteristics essential for thermal stability are likely to be more discernible [6]. Family pairs with habitat temperature differences <5°C were not considered significant and such terms were not included in the summations.

The temperatures within each family were randomly permuted within each family and family scores and overall ratios between the number of families with positive and negative trends calculated for the random datasets. This procedure was repeated 1000 times and averages and standard deviations calculated from the distribution of these results. The significance of an observed ratio for a given property was evaluated as the number of standard deviations above or below the average resulting from temperature shuffling (Z-score). A value of 2.0 (or greater) would indicate 95% (or greater) confidence in the statistical trend.

Surface contributions

The computer program ASC [26] was used to calculate analytic accessible surfaces and their atomic composition for all examined proteins. The default probe size for all calculations was 1.4 Å. The program generates lists of atoms and their accessible surface. Scripts written in Perl were used to cluster these lists and calculate the total and fractional contribution of different residue and atom types, polar (oxygen and nitrogen) and nonpolar (carbon and sulfur) atoms, and mainchain atoms versus sidechain atoms. Square temperature weighted trends within each family were calculated using equation 1 for each property.

Hydrogen bonds and salt links

Hydrogen bonds were calculated with the computer program HBPLUS [27] using the generally recommended [27,28] defaults for angular and distance constraints between donor (D), hydrogen (H), and acceptor (A) atoms and the atom covalently bound to A (AA). The program first generates hydrogens for all residues. Afterwards, the following constraints are applied: minimum angles D-H-A and D-A-AA, 90°; maximum distances between D-A, 3.9 Å and between H-A, 2.5 Å. Salt links were calculated using an ICM [9,21] script; they included all oppositely charged sidechain atom pairs not considered hydrogen bonded with a maximum separation of 4.0 Å following the proposal of Barlow and Thornton [29].

Packing density and protein cavities

The computer program ICM [9,21] was used to calculate packing densities and protein cavities. All default parameters of ICM were accepted unless otherwise noted. The algorithm first generates hydrogen atoms in their standard positions and then calculates analytically the surface touched by a sphere of a given radius rolling over the entire internal and external surface of the protein [30]. The resulting 'contact' surface is then allocated to the external protein surface and to buried cavities whose volumes and surfaces are also determined. Water was excluded from all calculations, as the reliability of the number and placement of water molecules is highly dependent on the X-ray crystallographic quality (resolution and refinement procedures) of every individual tertiary structure. Probe sphere radii (1.2–1.3 Å) for the different calculations were chosen following the suggestions of Hubbard and Argos [8] who found a probe radius range resulting in greatest consistency for cavity identification amongst various structural determinations of the same protein. Fractional cavity volumes are found by dividing the total cavity volume by the molecular volume, including those of subunits in the multimers.

The packing density of a protein is only loosely related to the number and size of cavities. Interstitial regions between the van der Waals' spheres of the protein atoms are considered as cavities only if the probe sphere mentioned above fits into this free space. Packing density which avoids the cavity limitations is calculated as the ratio between the volume enclosed by the contact surface of the protein and the volume occupied by the protein atoms, each assigned an appropriate van der Waals' radius [9,21]. Overlapping volume of covalently bonded atoms was taken only once.

Secondary structural characteristics

Secondary structural assignments for all protein residues were determined according to the automated computer algorithm STRIDE [31]. This assured consistency amongst all proteins and avoided biases due to different interpretations of the crystallographers. The secondary structural statistics were based on the pairwise sequence alignments amongst family members. A C implementation of the Needleman and Wunsch [32] algorithm was used to calculate these alignments with the optimal parameter settings found by Vogt *et al.* [20].

Acknowledgements

The authors are grateful for helpful discussions with Simon Hubbard and Stefanie Woell.

References

1. Gupta,M. (1993). *Thermostability of Enzymes*. Springer, Berlin.
2. Russell, R.J.M. & Taylor, G.L. (1995). Engineering thermostability: lessons from thermophilic proteins. *Curr. Opin. Biotechnol.* **6**, 370–374.
3. Querol, E., Perez-Pons, J.A. & Mozo-Villarias, A. (1996). Analysis of protein conformational characteristics related to thermostability. *Protein Eng.* **9**, 265–271.
4. Vieille, C. & Zeikus, J. (1996). Thermoenzymes: identifying molecular determinants of protein structural and functional stability. *Trends Biotechnol.* **14**, 183–191.
5. Perutz, M. & Raidt, H. (1975). Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature* **255**, 256–259.
6. Argos, P., Rossmann M.G., Grau, U.M., Zuber, H., Frank, G. & Tratschin, J.D. (1979). Thermal stability and protein structure. *Biochemistry* **25**, 5698–5703.
7. Bernstein, F.C., *et al.*, & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
8. Hubbard, S.J. & Argos, P. (1995). Detection of internal cavities in globular proteins. *Protein Eng.* **8**, 1011–1015.
9. Abagyan, R.A., Totrov, M. & Kuznetsov, D. (1994). ICM – a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488–506.
10. Fersht, A. & Serrano, L. (1993). Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* **3**, 75–83.
11. Pace, N.C., Laurents, D.V. & Erikson, R.E. (1992). Urea denaturation of barnase: pH dependence and characterization of the unfolded state. *Biochemistry* **31**, 2728–2734.
12. Yang, A.S. & Honig, B. (1993). On the pH dependence of protein stability. *J. Mol. Biol.* **231**, 459–474.
13. Branden, C. & Tooze, J. (1991). *Introduction to Protein Structure*. Garland Publishing Inc, New York.
14. Menéndez-Arias, L. & Argos, P. (1989). Engineering protein thermal stability: sequence statistics point to residue substitutions in alpha-helices. *J. Mol. Biol.* **206**, 397–406.
15. Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J.N. (1995). 2.0 Å structure of indols-3-glycerol phosphate synthase from the hyperthermophile S*ulfolobus solfataricus*: possible determinants of protein stability. *Structure* **3**, 1295–1306.
16. Korndörfer, I., Steipe, B., Huber, R., Tomschy, A. & Jaenicke, R. (1995). The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. *J. Mol. Biol.* **246**, 511–521.
17. Yip, K.S., *et al.*, & Consalvi, V. (1995). The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**, 1147–1158.
18. Salminen, T., Teplyakov, A., Karikare J., Cooperman, B.D., Lahti, R. & Goldman, A. (1996). An unusual route to thermal stability disclosed by the comparison of *Thermus thermophilus* and *Escherichia coli* inorganic pyrophosphatases. *Protein Sci.* **5**, 1014–1025.
19. Macedo-Ribeiro, S., Darimont, B., Sterner, R. & Huber, R. (1996). Small structural changes account for the high thermostability of 1[4Fe-4S] ferredoxin from the hyperthermophilic bacterium *Thermotoga maritima*. *Structure* **4**, 1291–1301.
20. Vogt, G., Etzold, T. & Argos, P. (1995). An assessment of amino acid exchange matrices in aligning protein sequences: the twilight zone revisited. *J. Mol. Biol.* **249**, 816–831.
21. Molsoft, LLC. (1995). ICM 2.0 program manual. World Wide Web URL: http://molsoft.com.
22. Krieg, N.R. & Holt, J.G. (1984). *Bergey's Manual of Systematic Bacteriology*. William and William, Baltimore, MD.
23. Barnett, J., Payne, R., & Yarrow, D. (1983). *Yeasts: Characteristics and Identification*. Cambridge University Press, Cambridge.
24. Herbert, R.A. & Sharp, R.J. (1992). *Molecular Biology and Biotechnology of Extremophiles*. Blackie & Son Ltd, New York.
25. Iny, D., Pinsky, A. & Malovani, H. (1993). The effect of cations on the thermophilic character of alkaline phosphatase from *Thermoactinomyces vulgaris*. *Biochem. Mol. Biol. Int.* **29**, 729–737.
26. Eisenhaber, F. & Argos, P. (1993). Improved strategy in analytic surface calculation for molecular systems: handling of singularities and computational efficiency. *J. Comput. Chem.* **14**, 1272–1280.
27. McDonald, I. & Thornton, J. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
28. Baker, E. & Hubbard, R. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44**, 97–179.
29. Barlow, D. & Thornton J. (1983). Ion-pairs in proteins. *J. Mol. Biol.* **168**, 867–885.
30. Totrov, M.M. & Abagyan, R.A. (1995). The contour-buildup algorithm to calculate the analytical molecular surface. *J. Struct. Biol.* **115**, 1–6.
31. Frishman, D. & Argos, P. (1995). Knowledge-based secondary structure assignment. *Proteins* **23**, 566–579.
32. Needleman, S. & Wunsch, C. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453.